Anal Bioanal Chem (2011) 399:2719–2730
DOI 10.1007/s00216-010-4282-7

ORIGINAL PAPER

Innovative methodology to transfer conventional GC-MS heroin profiling to UHPLC-MS/MS

B. Debrus · J. Broséus · D. Guillarme · P. Lebrun · P. Hubert · J.-L. Veuthey · P. Esseiva · S. Rudaz

Received: 3 August 2010 / Revised: 24 September 2010 / Accepted: 29 September 2010 / Published online: 17 October 2010

Abstract Nowadays, in forensic laboratories, heroin profiling is frequently carried out by gas chromatography coupled with mass spectrometry (GC-MS). This analytical technique is well established, provides good sensitivity and reproducibility, and allows the use of large databases. Despite those benefits, recently introduced analytical techniques, such as ultra-high-pressure liquid chromatography (UHPLC), could offer better chromatographic performance, which needs to be considered to increase the analysis throughput for heroin profiling. With the latter, chromatographic conditions were optimized through commercial modeling software and two atmospheric pressure ionization sources were evaluated. Data obtained from UHPLC–MS/MS were thus transferred, thanks to mathematical models to mimic GC-MS data. A calibration and a validation set of representative heroin samples were selected among the database to establish a transfer methodology and assess the models' abilities to transfer using principal component analysis and hierarchical classification analysis. These abilities were evaluated by computing the frequency of successful classification of UHPLC–MS/MS data among GC-MS database. Seven mathematical models were tested to adjust UHPLC–MS/MS data to GC-MS data. A simplified mathematical model was finally selected and offered a frequency of successful transfer equal to 95%.

Keywords Heroin profiling · Principal component analysis · Hierarchical classification analysis · UHPLC–MS/MS · Multiple linear regression · Transfer

Published in the special issue on *Advances in Analytical Mass Spectrometry* with Guest Editor Maria Careri.

B. Debrus (✉) · P. Lebrun · P. Hubert
Laboratory of Analytical Chemistry, Department of Pharmacy, CIRM, University of Liège,
Avenue de l'Hôpital 1, B36, Tour4, +2,
4000 Liège, Belgium
e-mail: b.debrus@ulg.ac.be

J. Broséus · P. Esseiva
School of Criminal Sciences, University of Lausanne,
UNIL-Sorge, Bâtiment Batochime 6407,
1015 Lausanne, Switzerland

D. Guillarme · J.-L. Veuthey · S. Rudaz (✉)
School of Pharmaceutical Sciences, University of Geneva, University of Lausanne,
Boulevard d'Yvoy 20,
1211 Geneva 4, Switzerland
e-mail: serge.rudaz@unige.ch

Introduction

The profiling of heroin follows several objectives. One of them is the establishment of links between samples that have been seized at different locations in order to highlight information that may be used to decipher drug trafficking organization and distribution pattern for supporting law enforcement investigation. The improvement and quickening of drug profiling processes remains a constant preoccupation to provide timely results necessary for an effective battle against drug manufacturers and drug dealers. Heroin (diacetylmorphine), is a semi-synthetic compound derived from morphine, which is extracted from *Papaver somniferum*. During the extraction process, many compounds are simultaneously extracted, which can also be transformed during the acetylation of morphine. Therefore, proportions of co-

extracted molecules vary depending on the method of cultivation and the acetylation process. Subsequent cuts (solid dilution) do not affect the relative proportions of heroin and co-extracted molecules, and a comparison of seized drugs is possible to highlight traffic networks for providing both tactical and strategic intelligence [1–8]. Today, the profiling of heroin is generally performed by gas chromatography coupled with mass detection (GC-MS) [9]. To increase the volatility of those molecules for GC-MS analysis, a derivatization step with *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) is mandatory. This derivatization step takes about an hour, allowing compounds to be analyzed by GC-MS in <25 min (see Fig. 1). An existing GC-MS database combines results of 859 samples of heroin seized in a French-speaking part of Switzerland during the last 2 years (2008–2009). In this database, peak areas of seven compounds are compiled: meconine (MEC), acetylcodeine, acetylthebaol, 6-monoacetylmorphine (6-MAM), heroin, papaverine, and noscapine (NOS). GC-MS profiling methodology involves, among others, the standardization of data and the use of principal component analysis (PCA). Despite the quality of profiling results obtained by Esseiva et al. [2, 9] and Dujourdy et al. [3], the complete GC-MS procedure is long and some recent analytical techniques may be envisaged to conduct heroin profiling analysis.

In these last years, liquid chromatography (LC) has strongly evolved and numerous manufacturers offer new equipment and chromatographic columns allowing very fast and/or highly efficient separations [10–13]. In our opinion, those benefits have to be considered for heroin profiling. In this paper, ultra-high-pressure liquid chromatography (UHPLC) was selected to investigate a possible transfer of the conventional heroin profiling method toward alternative, rapid, and selective LC determinations. Besides, the applicability of UHPLC–MS/MS for heroin profiling was already discussed in the work of Lurie and Toske [14]. Today, it is admitted that to maintain a harmonized database based on GC-MS determination, it is compulsory to adopt strictly similar analytical conditions (i.e., the same brand of GC-MS, liner, column, etc.). This approach's capability is thus somehow limited when one needs to renew an old instrument or replace it with a more efficient analytical technique (e.g., fast separation). Because relative peak intensities are often different, it is then necessary to create a new database and consequently reset the memory of the previous knowledge. In order to overcome this problem, an innovative data treatment methodology was proposed. In the present study, the major challenge was to demonstrate the possibility of feeding a common conventional GC-MS database with data coming from a different analytical technique (i.e., UHPLC–MS/MS).

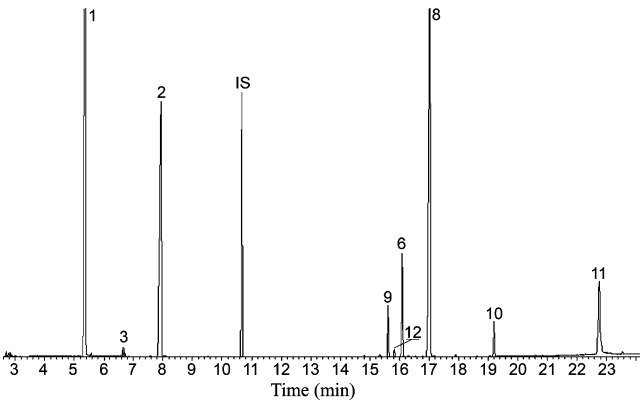

Fig. 1 Total ion chromatogram of a heroin sample obtained by GC-MS. Column: non-polar DB-5 ms capillary column (30-m length, 0.25-mm i.d., and 0.25-μm film thickness), the initial oven temperature was set at 150 °C, first raised to 250 °C (at 8 °C/min) and then to 320 °C (at 6 °C/min), helium flow=1 mL/min. Peak identity (numbered in accordance with UHPLC elution order): *1* paracetamol, *2* caffeine, *3* meconine, *6* 6-monoacetylmorphine, *8* diacetylmorphine, *9* acetylcodeine, *10* papaverine, *11* noscapine, *12* acetylthebaol, *IS* internal standard

Experimental

All samples were obtained in street-seized powdered form. Samples were seized by the Swiss police in 2008 and 2009.

Chemicals and reagents

For UHPLC–MS/MS experiments, formic acid and methanol (MeOH) were of ULC/MS grade and purchased from Biosolve (Valkenswaard, the Netherlands). Ammonium hydroxide was provided by Sigma-Aldrich (Buchs, Switzerland). Water was obtained from a Milli-Q Water Purification System from Millipore (Bedford, MA, USA). Ammonia buffer, 10 mM, was prepared with an adapted volume of ammonium hydroxide and the pH adjusted to 8.0, 8.5, and 9.0 with formic acid. pH was measured with a Metrohm pH meter (Herisau, Switzerland), and the prepared buffer had a buffer capacity higher than 5 mM/pH unit. For GC-MS experiments, chloroform, pyridine, heneicosane, and MSTFA were purchased from Fluka (Sigma-Aldrich).

GC-MS instrumentation

An Agilent Technologies 6890A gas chromatograph interfaced with an Agilent Technologies 5975C mass selective detector was used for GC-MS analyses. Analytes were separated on a non-polar DB-5 ms capillary column (30-m length, 0.25-mm i.d., and 0.25-μm film thickness, J&W

Scientific, Agilent Technologies, Palo Alto, CA, USA). The initial oven temperature was 150 °C, first raised to 250 °C (at 8 °C/min) and then to 320 °C (at 6 °C/min) for a total run of 24.17 min. Two microliters of each sample was injected using helium as carrier gas (constant flow mode, 1 mL/min). Injections were carried out in split mode using a split/splitless liner packed with glass wool (Agilent Technologies, no. 5183-4711) with a split ratio of 1:50. Applied temperatures were: 250 °C for injector, 280 °C for transfer line, 230 °C for ion source, and 150 °C for quadrupole. Data were acquired in the full scan mode (30–450 *m*/*z* mass range) with a sampling rate of 3 (1.77 scans per second). Data were then analyzed using MSD Enhanced ChemStation v. D.02.00.275 (Agilent Technologies). For the determination of the chemical signature of heroin samples, peak areas of targeted ions were studied (see Table 1).

UHPLC–MS/MS instrumentation

UHPLC–MS/MS experiments were performed on a Waters Acquity UPLC system (Milford, MA, USA) hyphenated with a triple quadrupole TQD mass spectrometer from Waters (Acquity TQD detector). The UPLC instrument included a binary pumping system with a maximum flow rate of 2 mL/min, an autosampler with an injection loop volume of 2 μL used in full loop conditions, a UV–Vis programmable detector, and a column manager that included a column oven set at 30 or 50 °C. Separation was carried out on Acquity BEH C18 50×2.1 mm, 1.7 μm, and Acquity BEH Shield RP18 50×2.1 mm, 1.7 μm, both provided by Waters. The following solvent system was considered: A = ammonium buffer 10 mM; B = methanol or acetonitrile. Optimal chromatographic conditions were found, thanks to HPLC modeling software (Osiris 4.1.1.2, Datalys, Grenoble, France). For this purpose, analytes were individually injected at 500 μL/min using two gradient runs that differ in slope, namely, 5% to 95% B linear gradient in 14.4 and 4.8 min, respectively. Optimal chromatographic conditions at various pH, mobile phases, and stationary phases were determined for t_{min} (i.e., the minimum acceptable elution time) equal to 0.36 min and t_{max} (i.e., the maximum acceptable elution time) of 3.8 min.

The TQD operated at single mass resolution of *m*/*z* 0.7 full width at half-maximum and possesses an upper mass limit of *m*/*z* 2,000. The dual atmospheric pressure chemical ionization (APCI)/electrospray ionization (ESI) source was used in the ESI positive mode and ionization parameters, cone voltages, and collision energies optimized by infusing each compound (1 μg/mL) in 30:70 MeOH/water at a flow rate of 500 μL/min. Optimal cone voltages and collision energies values were summarized in Table 2. The capillary voltage and the source extractor voltage were set at 3,000 and 3 V, respectively. The source temperature was maintained at 150 °C, the desolvatation gas temperature and flow at 400 °C and 800 L/h, respectively, and the cone gas flow at 50 L/h. MS/MS detection was carried out in the multiple reaction monitoring (MRM) mode and the transitions also indicated in Table 2.

The collision gas flow was set at 0.2 mL/min of argon and entrance and exit potentials, respectively adjusted to 1 and 0.5 V. Some experiments were also carried out with a dual APCI/atmospheric pressure photo-ionization (APPI) source used in the positive APPI mode. The parameters for APPI source were the following: the repeller and the source extractor voltage were set at 1,000 and 3 V, respectively. The source temperature was maintained at 120 °C, the APCI/APPI probe temperature and flow at 550 °C and 300 L/h, respectively, and the cone gas flow at 10 L/h. Data acquisition, data handling, and instrument control were performed by Masslynx v4.1 Software.

Sample preparation

For GC-MS, three replicates of each heroin sample (8 mg each) were weighed with a MX5 microbalance (Mettler Toledo, Columbus, OH, USA). Each replicate was extracted in 500 μL of a solution containing 1 mg/mL of heneicosane (Internal Standard) in a mixture of chloroform and pyridine (5:1, *v*/*v*). Derivatization was performed by adding 100 μL of MSFTA followed by heating at 80 °C for 60 min. Finally, an aliquot was placed into a vial for injection in the GC system. Each replicate was injected once and the mean peak area (i.e., mean on the three replicates) was considered for calculation.

For UHPLC–MS/MS, a stock solution at 1 mg/mL was prepared by dissolving 1 mg of each heroin sample (weighted on a Mettler Toledo MX5 microbalance) in 1 mL of MeOH. Then, an aliquot of 100 μL was diluted with 200 μL of MeOH and 700 μL of water for injection in

Table 1 List of the targeted ions for each compound extracted for the determination of the chemical signature of heroin samples in GC-MS

No.	Compound	t_R (min)	Target ion (*m*/*z*)
3	Meconine	6.7	194.1
9	Acetylcodeine	15.6	341.2
12	Acetylthebaol	15.8	254.1
6	6-MAM	16.1	399.3
8	Diacetylmorphine	17.0	369.2
10	Papaverine	19.2	338.2
11	Noscapine	22.8	220.1

Peak numbering was done in accordance with UHPLC elution order

Table 2 Molecular weights, MS/MS transitions, dwell times, optimal cone voltages, and collision energies for the investigated compounds during the profiling of unknown heroin samples with UHPLC–MS/MS

No.	Compound	t_R (min)	MW	MS/MS transitions	Dwell time (ms)	Cone voltage (V)	Collision energy (V)
3	Meconine	0.83	194.0	195.0>162.0	40	30	20
				195.0>180.0	40	30	20
6	6-MAM	1.67	327.3	328.3>165.2	5	40	40
				328.3>211.3	5	40	25
8	Diacetylmorphine	2.2	369.4	370.4>165.3	5	40	50
				370.4>268.3	5	40	30
9	Acetylcodeine	2.38	341.3	342.3>165.3	5	40	50
				342.3>225.3	5	40	25
10	Papaverine	2.58	339.3	340.3>171.0	5	45	40
				340.3>202.5	5	45	25
11	Noscapine	3.21	413.3	414.3>220.3	5	35	20
				414.3>353.3	5	35	25
12	Acetylthebaol	3.84	296.0	297.0>223.3	50	20	10
				297.0>255.3	50	20	10

the UHPLC system. Heroin samples were freshly diluted in the appropriate solution just before analyses in order to avoid the hydrolysis of alkaloids.

Transfer methodology

To introduce the problematic encountered in the present study, a simple example can be taken, namely, the comparison of peak areas obtained by two analytical techniques. If standard reference is available and if no formal quantitative determination is required, the transfer methodology can be simply based on the comparison of both calibration curves (i.e., the relationship between concentration and analytical response in a given concentration range). But when standard reference is not available, it is mandatory to establish the mathematical relationship linking peak areas coming from both techniques (i.e., transfer rules) by the use of various unknown concentration samples. The case study dealing with the transfer of conventional GC-MS heroin profiling to UHPLC–MS/MS was very similar to this scenario, but seven compounds were involved. Thus, seven mathematical equations had to be established to transfer peak areas of the seven compounds and introduce heroin profiles corresponding to each investigated sample in the GC-MS database using peak areas coming from UHPLC–MS/MS. The developed methodology allowed converting UHPLC–MS/MS peak areas in "GC-like" data for their implementation in the original GC-MS database. The following sections describe the steps to successfully achieve this transfer and evaluate its validity. The corresponding flowchart is depicted in Fig. 2.

GC-MS data

The GC-MS database consists of the aforementioned seven compound peak areas for 859 heroin samples. For each heroin sample, peak areas were normalized by dividing them by their sum to balance the weight of each variable and compensate the variability of MS data intensity. Subsequently, unit variance scaling was performed for each compound; relative areas were centered and reduced by subtracting the mean and dividing by the standard deviation. This variance scaling step is mandatory to find reliable projections.

UHPLC–MS/MS data

For this step, 58 heroin samples were selected by a stratified random sampling. The strata were the clusters coming from GC-MS sample classifications obtained by a previous work [9]. This sample set was divided in a calibration set (36 samples) allowing to fit transfer models and a validation set (22 samples) used to estimate predictive abilities of the transfer models. These heroin samples were analyzed by UHPLC–MS/MS and peak areas were normalized, centered, and reduced in the same way as GC-MS data (see "GC-MS data"). The GC-MS database was then divided into three subsets: the GC-MS calibration set data (i.e., the 36 heroin samples included in the calibration set, analyzed by GC-MS), the GC-MS validation set data (i.e., the 22 heroin samples, included in the validation set, analyzed by GC-MS), and the reduced GC-MS database (containing the data of the 801 heroin samples analyzed by GC-MS which are neither in the calibration nor in the validation set).

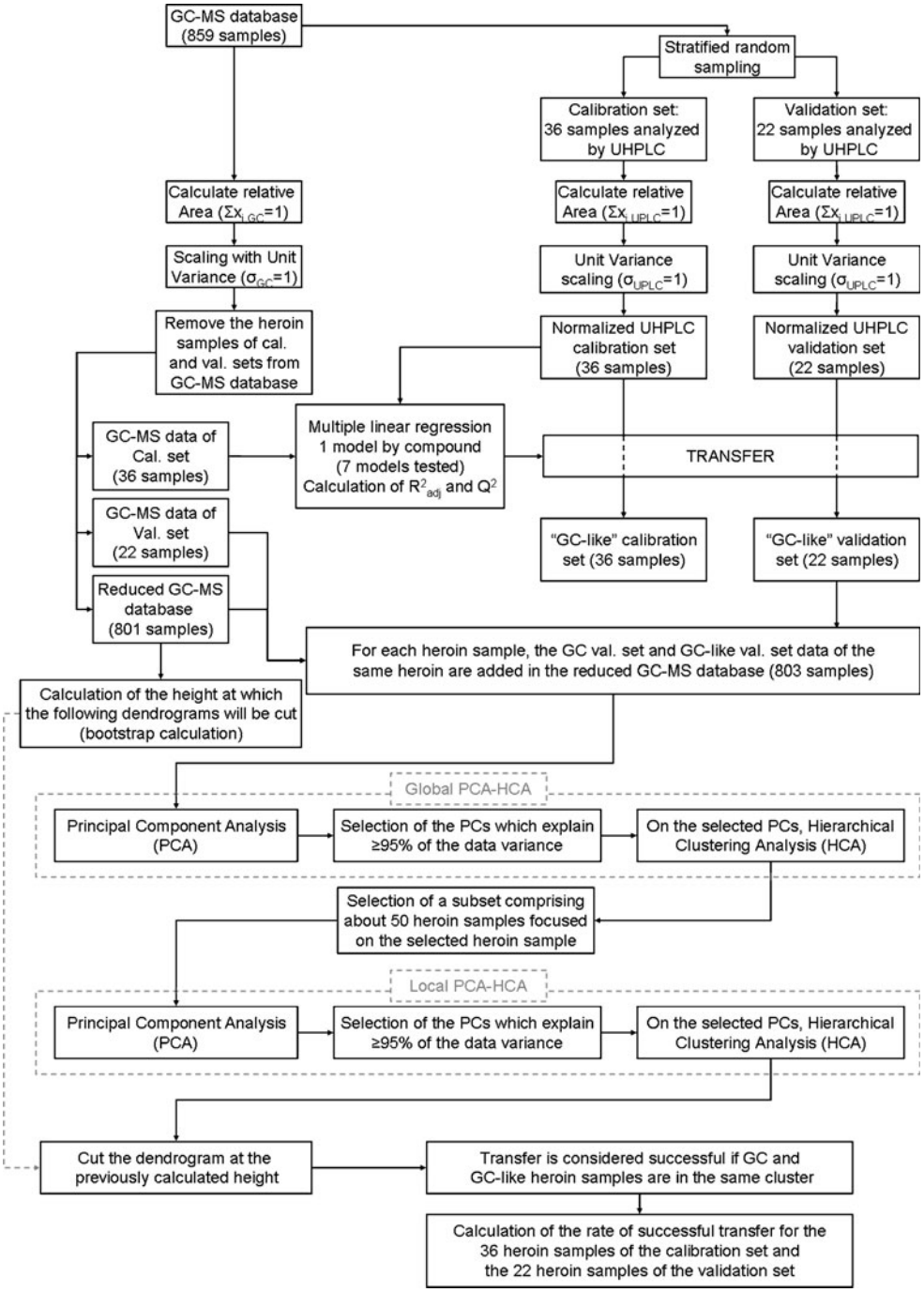

Fig. 2 Flowchart of the methodology used in the present study

Transfer

As the objective was to predict peak areas that would have been obtained by GC-MS from those obtained in UHPLC–MS/MS, model responses were GC-MS peak areas. For each heroin sample, seven mathematical models were employed to transform each peak area. For instance, noscapine area obtained by UHPLC–MS/MS (NOS_{UHPLC}) was transformed, using a selected transfer model, into a "GC-like" noscapine area ($NOS_{GC\text{-}like}$) which needs to be as close as possible to the noscapine area obtained in GC-MS (NOS_{GC}).

Mathematical models

Numerous mathematical models were considered, including the simplest mathematical approaches, such as linear, quadratic, or cubic models, but also more complex such

Table 3 Mathematical models tested for the transfer to GC-MS database (noscapine as example)

Name	Model
Linear	$NOS_{GC} = \beta_1 \cdot NOS_{UHPLC} + \beta_0$
Quadratic	$NOS_{GC} = \beta_2 \cdot NOS_{UHPLC}^2 + \beta_1 \cdot NOS_{UHPLC} + \beta_0$
Cubic	$NOS_{GC} = \beta_3 \cdot NOS_{UHPLC}^3 + \beta_2 \cdot NOS_{UHPLC}^2 + \beta_1 \cdot NOS_{UHPLC} + \beta_0$
Poly-linear	$NOS_{GC} = \beta_1 \cdot NOS_{UHPLC} + \beta_4 \cdot MEC_{UHPLC} + \beta_5 \cdot AC_{UHPLC} + \beta_6 \cdot AcTB_{UHPLC} + \beta_7 \cdot MAM_{UHPLC} + \beta_8 \cdot DAM_{UHPLC} + \beta_9 \cdot PAP_{UHPLC} + \beta_0$
Poly-quadratic	$NOS_{GC} = \beta_1 \cdot NOS_{UHPLC} + \beta_4 \cdot MEC_{UHPLC} + \beta_5 \cdot AC_{UHPLC} + \beta_6 \cdot AcTB_{UHPLC} + \beta_7 \cdot MAM_{UHPLC} + \beta_8 \cdot DAM_{UHPLC} + \beta_9 \cdot PAP_{UHPLC} + \beta_{10} \cdot NOS_{UHPLC}^2 + \beta_{12} \cdot MEC_{UHPLC}^2 + \beta_{13} \cdot AC_{UHPLC}^2 + \beta_{14} \cdot AcTB_{UHPLC}^2 + \beta_{15} \cdot MAM_{UHPLC}^2 + \beta_{16} \cdot DAM_{UHPLC}^2 + \beta_{17} \cdot PAP_{UHPLC}^2 + \beta_0$
Complete	$NOS_{GC} = \beta_1 \cdot NOS_{UHPLC} + \beta_4 \cdot MEC_{UHPLC} + \beta_5 \cdot AC_{UHPLC} + \beta_6 \cdot AcTB_{UHPLC} + \beta_7 \cdot MAM_{UHPLC} + \beta_8 \cdot DAM_{UHPLC} + \beta_9 \cdot PAP_{UHPLC} + \beta_2 \cdot NOS_{UHPLC}^2 + \beta_{10} \cdot MEC_{UHPLC}^2 + \beta_{11} \cdot AC_{UHPLC}^2 + \beta_{12} \cdot AcTB_{UHPLC}^2 + \beta_{13} \cdot MAM_{UHPLC}^2 + \beta_{14} \cdot DAM_{UHPLC}^2 + \beta_{15} \cdot PAP_{UHPLC}^2 + \beta_{16} \cdot NOS_{UHPLC} \cdot MEC_{UHPLC} + \beta_{17} \cdot NOS_{UHPLC} \cdot MEC_{UHPLC}^2 + \beta_{18} \cdot NOS_{UHPLC}^2 \cdot MEC_{UHPLC} + \beta_{19} \cdot NOS_{UHPLC} \cdot AC_{UHPLC} + \beta_{20} \cdot NOS_{UHPLC} \cdot AC_{UHPLC}^2 + \beta_{21} \cdot NOS_{UHPLC}^2 \cdot AC_{UHPLC} + \beta_{22} \cdot NOS_{UHPLC} \cdot AcTB_{UHPLC} + \beta_{23} \cdot NOS_{UHPLC} \cdot AcTB_{UHPLC}^2 + \beta_{24} \cdot NOS_{UHPLC}^2 \cdot AcTB_{UHPLC} + \beta_{25} \cdot NOS_{UHPLC} \cdot MAM_{UHPLC} + \beta_{26} \cdot NOS_{UHPLC} \cdot MAM_{UHPLC}^2 + \beta_{27} \cdot NOS_{UHPLC}^2 \cdot MAM_{UHPLC} + \beta_{28} \cdot NOS_{UHPLC} \cdot DAM_{UHPLC} + \beta_{29} \cdot NOS_{UHPLC} \cdot DAM_{UHPLC}^2 + \beta_{30} \cdot NOS_{UHPLC}^2 \cdot DAM_{UHPLC} + \beta_{31} \cdot NOS_{UHPLC} \cdot PAP_{UHPLC} + \beta_{32} \cdot NOS_{UHPLC} \cdot PAP_{UHPLC}^2 + \beta_{33} \cdot NOS_{UHPLC}^2 \cdot PAP_{UHPLC} + \beta_0$
Simplified	Terms of the complete model selected with the Pareto analysis

Mathematical terms were labeled according to the compound, with subscripts corresponding to the analytical technique

as poly-linear or poly-quadratic models, as presented in Table 3 where, for the sake of clarity, only the example of one analyte (noscapine) is presented.

Mathematical models can be classified into two sets. The first one includes simple models (i.e., linear, quadratic, and cubic models) which only take into account the predicted variable (e.g., only NOS_{UHPLC} and higher order effects were used to predict NOS_{GC}). The second set contains the models which use all the effects (i.e., the mathematical terms) to predict one variable (e.g., all the UHPLC–MS/MS peak areas were used to predict NOS_{GC}). The poly-linear model takes into account areas (neither squared nor cubed areas) of all compounds to predict one of them. The poly-quadratic model also considers the influence of the squared areas of other compounds. The complete model included the cross-terms (e.g., $NOS_{UHPLC} \cdot MEC_{UHPLC}$) and the ones containing squared terms (e.g. $NOS_{UHPLC} \cdot MEC_{UHPLC}^2$ and $NOS_{UHPLC}^2 \cdot MEC_{UHPLC}$). This model was named "complete" because it consists of 34 mathematical terms for a maximum of 35 since one degree of freedom was necessary to the error estimation, ε. To simplify the complete models, Pareto analyses were carried out on each complete model. In this analysis, coefficients of each term (β_0, $\beta_1 \ldots \beta_{33}$) were squared and divided by the sum of the squared β to obtain the relative contribution of each term (i.e., relative β). The terms were then sorted according to their influence (i.e., the weight) to the goodness of fit. As depicted in Fig. 3 for acetylthebaol, many mathematical terms had a relatively small contribution and were removed from the mathematical model to avoid overfitting. Finally, the most contributive mathematical terms were selected to create the simplified models with the following criteria: Contributive terms were selected until the cumulative response explanation reached 80%. In addition, poly-linear, poly-quadratic, and complete modelings were

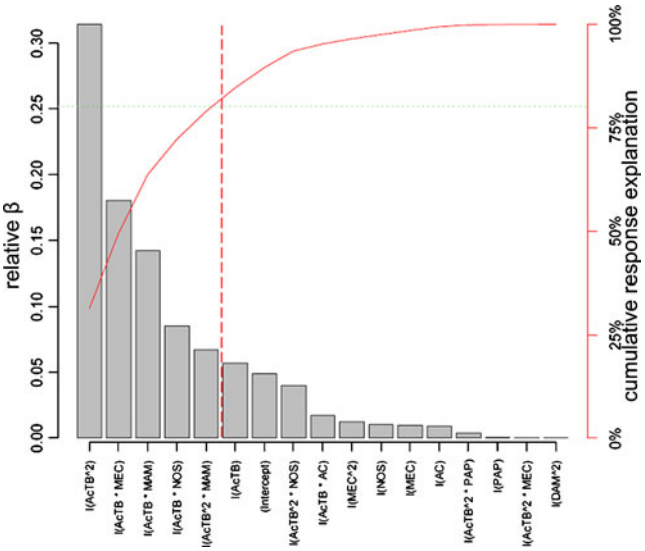

Fig. 3 Example of Pareto analysis for the complete model with acetylthebaol. The *red curve* depicts the cumulative response explanation. The *red dashed vertical line* represents the limit of selection. In this example, the most explanatory terms (*left* of the *red vertical line*) were selected to create the simplified model

achieved through a stepwise regression that selected the terms, maximizing the adjusted determination coefficients (adjusted R^2) [15]. See the Appendix for the complete equations of the seven simplified models.

Multiple linear regressions, model adjustment, and evaluation

The adjustments of mathematical models were carried out by performing multiple linear regressions (MLR) of the UHPLC–MS/MS calibration set against the GC-MS calibration set. To evaluate the quality of MLRs, the mean adjusted R^2 were calculated (i.e., for a given tested model, the mean adjusted R^2 obtained for each compound). Furthermore, the prediction coefficients (Q^2) were also approximated using a leave-n-out cross-validation calculation (with n=30% of the calibration set size). Thus, 11 samples were randomly removed from the calibration set to create a temporary set. This set was considered to adjust the multiple linear models, while the remaining calibration samples were used to estimate Q^2. The procedure was repeated 250 times for each model in order to obtain a robust estimation of Q^2. Finally, using the adjusted models, UHPLC–MS/MS calibration set and UHPLC–MS/MS validation set were respectively transformed into the "GC-like" calibration set and the "GC-like" validation set (i.e., back-calculated GC-MS peak areas).

Data clustering

During the heroin profiling process, samples are generally classified to highlight connections between samples. In the present study, the clustering procedure consisted in a sequential unsupervised multiway data analysis, as schematized in Fig. 4. PCA was initially carried out to select the principal components (PCs) explaining at least 95% of the initial data variance. This step mainly consisted in data cleaning since the non-selected PCs mainly correspond to non-significant variability. Then, a hierarchical clustering analysis (HCA) was performed using the sample coordinates in the space of the selected PCs as described by Stella et al. [16] and Boccard et al. [17]. HCA was accomplished with the Euclidean distances and the Ward algorithm [18]. This corresponds to a global PCA–HCA, carried out on the reduced GC-MS database within which GC-MS data and "GC-like" data of one tested heroin sample were added (803 samples).

In the resulting dendrogram, the last step was the selection of a subset comprising at least 50 heroin samples in the immediate neighborhood of the tested "GC-like" sample. On this subset, a similar PCA–HCA was performed to refine the evaluation of the sample similarity in the selected cluster and thus increase the predictions accuracy. This corresponds to a local PCA–HCA, carried out on the most similar samples of the tested "GC-like" sample.

Significant clustering height calculation

To obtain a clustering after HCA, the dendrogram was cut at a given height. In the present study, this height was calculated using the freely available "pvclust" R-package so that the p values for each cluster were calculated by multiscale bootstrap resampling [19]. The reduced GC-MS database (801 samples) was used to estimate this cutting

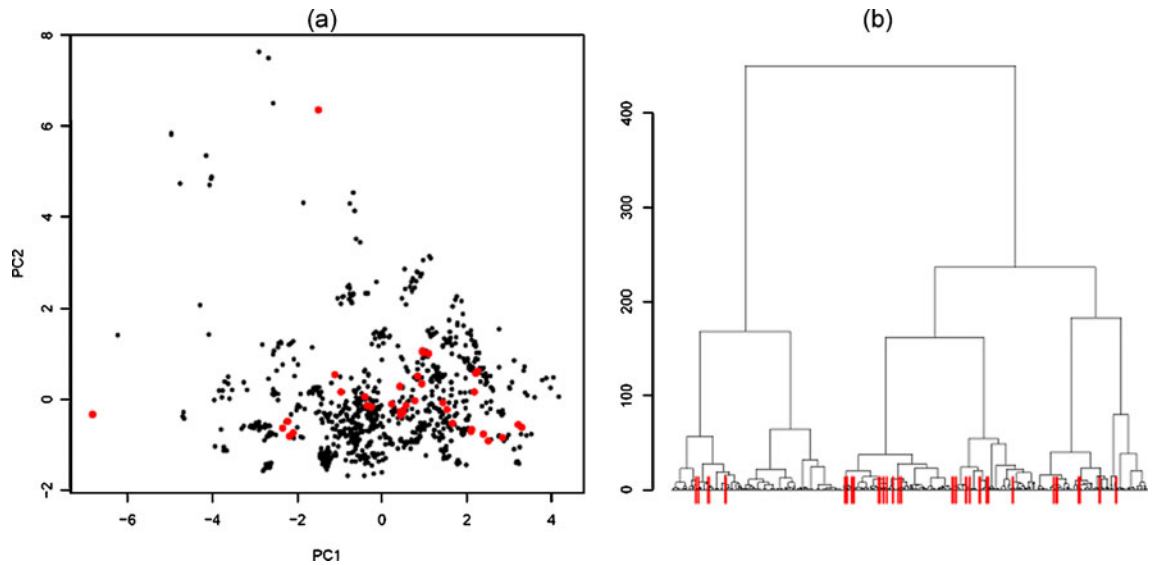

Fig. 4 Principal component analysis and hierarchical cluster analysis results: **a** PC1 vs. PC2 plot of 859 heroin samples. **b** Hierarchical cluster analysis performed using the selected PC coordinates. The 36 samples of the calibration set are displayed in *red*

height to get significant p values which indicated that clusterings were data-driven. This height was found to be equal to 20.0.

Evaluation of transfer model predictive abilities

To evaluate the quality (i.e., predictive abilities) of the transfer models, the frequency of correct classification was computed on the corresponding local PCA–HCA dendrogram cut at the height mentioned above. When the tested "GC-like" sample (i.e., a heroin analyzed by UHPLC–MS/MS after data transformation) was identified in the same final cluster as the GC-MS heroin sample in the local dendrogram, the transfer of heroin profile coming from UHPLC–MS/MS was considered successful. On this basis, the predictive ability of each considered mathematical model was quantified by counting the number of samples where GC and "GC-like" data were found in the same cluster, i.e., correctly transferred. Finally, this procedure was carried out for all heroin samples of the calibration and the validation sets, allowing for the determination of an individual frequency of successful transfer for each mathematical model under investigation.

Results and discussion

Twelve components, including the seven alkaloids used in heroin profiling, were detected during UHPLC–MS/MS analysis. The separation and MS signal were optimized to achieve good sensitivity, precision, and accuracy of the method (see Fig. 5 for a UHPLC–MS/MS chromatogram).

UHPLC–MS/MS conditions

In these last few years, LC has evolved to improve chromatographic performance. For this purpose, different strategies have recently been developed to minimize the analysis time while maintaining efficiency or to improve efficiency without sacrificing analysis time [20, 21]. Among them, the use of column packed with sub-2-μm particles becomes more popular, thanks to the recent introduction of new instruments able to withstand ultra-high pressures (up to 1,300 bar). The hyphenation with MS detection is now well established, but besides the separation aspect, the choice of an adequate ionization process is of prime importance in condensed phase.

Determination of MS/MS conditions

In the present study, two different ionization sources were initially evaluated, namely, ESI and APPI. The latter was

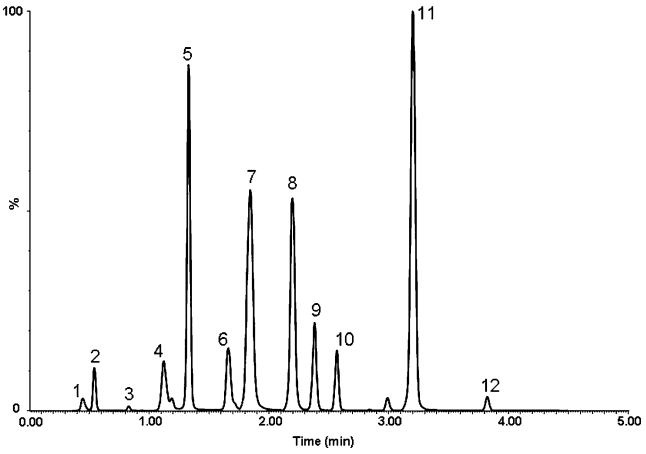

Fig. 5 BPI (base peak intensity) chromatogram of the UHPLC–MS/MS separation obtained after optimization of the chromatographic separation for the 12 compounds of interest. Column: Acquity BEH Shield RP18 50×2.1 mm, 1.7 μm, T=30 °C, F=500 μL/min, pH 9. Gradient from 27% to 69.5% MeOH in 3.48 min, followed by a step at 69.5% for 0.52 min. Peak identity: *1* paracetamol, *2* caffeine, *3* meconine, *4* morphine, *5* phenacetin, *6* 6-MAM, *7* procaine, *8* heroin, *9* acetylcodeine, *10* papaverine, *11* noscapine, *12* acetylthebaol

investigated because it is theoretically adapted to a wider range of analyte polarity [22] despite it being well known to provide slightly lower sensitivity than ESI. Various experimental conditions were first evaluated in APPI: without dopant, with acetone 5%, toluene 5%, and chlorobenzene 5% [23]. Dopant was added to enhance the ionization of less polar analytes through charge exchange with dopant radical cations. In the presence or absence of dopant, it was possible to detect heroin, impurities from opium, solid diluents, and impurities which arise from acetylation of opium alkaloids, except acetylthebaol, the most apolar compound, which was surprisingly not ionized with our APPI source.

In a second step, ESI was evaluated. After optimizing MS parameters, ionization of all compounds of interest, even acetylthebaol, was possible. The sensitivity (in terms of signal-to-noise ratio) was on average tenfold higher in ESI compared to APPI, as reported elsewhere [24]. It is, however, worth mentioning that ionization of acetylthebaol was critical in ESI and cone voltage (CV) should be carefully selected to avoid in-source fragmentation. Indeed, the peak height drops when increasing CV beyond 20 V, while the peak was not detected anymore at 50 V. Thus, optimal CV for acetylthebaol was set at 20 V.

Using optimal ESI settings reported in "UHPLC–MS/MS instrumentation", product ion scans were acquired with collision energies ranging between 10 and 60 eV, and the two most intense transitions were selected for each monitored precursor m/z. The optimal transitions as well as the corresponding collision energies were reported in

Table 2. Regarding optimal SRM transitions, the selection was made in agreement with data from the literature [25–27], while the explanation of fragmentation pathways is out of the scope of the present paper.

Optimization of UHPLC separation

To reach suitable chromatographic selectivity and avoid ion suppression or enhancement effects arising from the co-elution of several compounds, the mobile phase pH (i.e., pH of 8.0, 8.5, and 9), temperature (i.e., 30 and 50 °C), organic modifier (i.e., acetonitrile and methanol), and stationary phase (i.e., Acquity BEH C18 and Acquity BEH Shield RP18) were tuned during chromatographic method development with a 100 μg/mL heroin sample containing all analytes and further spiked with 1 μg/mL of paracetamol, caffeine, phenacetin, morphine, and procaine. For all these combinations of parameters, representing 24 set of conditions, two gradient experiments from 5% to 95% in 4.8 and 14.4 min were carried out at a flow rate of 500 μL/min. Then, data were computed in HPLC modeling software to find out the best separation.

A first remark concerns the choice of pH. It is generally well established that LC-MS separations of basic drugs, such as those investigated in the heroin profiling, should be carried out in acidic conditions to maximize sensitivity with fully ionized molecules. However, in these chromatographic conditions, poor chromatographic performance was obtained, leading to detrimental resolution. For this reason, basic pH was preferentially selected as it allows higher selectivity and retention, but also better sensitivity, as recently demonstrated with a mixture of drugs and metabolites eluted at pH 9 [28]. This behavior was attributed to a better desolvatation when analytes were eluted with a high proportion of organic solvent. In the present study, the pH has a strong influence on the separation quality because most of compounds possess pK_a close to the investigated pH. For this separation, the highest pH value was found to be beneficial, whatever the temperature, nature of mobile, and stationary phases. However, as the column lifetime can be reduced in too aggressive conditions (i.e., the Acquity BEH Shield RP18 can theoretically be used up to pH 11 because of ethylene bridges inside the silica matrix, which prevent its dissolution), more alkaline pH values were avoided.

Regarding the choice of mobile phase temperature, the interest of working at 50 vs. 30 °C was evaluated. As shown by Nguyen et al. [29], faster separations can be attained, with a reasonable backpressure in UHPLC at higher temperature. However, in the present study, selectivity was slightly lower at 50 vs. 30 °C, particularly for the separation of heroin (one of the largest peaks), acetylcodeine, and papaverine. Most importantly, the stability of the chromatographic support when simultaneously increasing temperature and pH becomes critical. Thus, the separation was preferentially carried out at 30 °C, at pH 9, to avoid such degradation issue.

Two organic modifiers (i.e., methanol and acetonitrile) were evaluated, and methanol provided the best selectivity, particularly for the critical separation of heroin/papaverine/acetylcodeine. In addition, because of the recent acetonitrile shortage, methanol appeared as the solvent of choice to limit experimental cost [30]. The only drawback associated with the use of methanol in UHPLC is its elevated viscosity, generating higher backpressure. However, as the flow rate was 500 μL/min and column length equals 50 mm, backpressure was always lower than 700 bar.

Finally, two columns from the same provider (Acquity BEH C18 and Acquity BEH Shield RP18) were compared to evaluate the selectivity between the 12 detected compounds (i.e., seven investigated alkaloids, two co-extracted alkaloids, and three solid diluents). A significant difference was observed between these two stationary phases. The C18 Shield material, containing an embedded carbamate group, appeared to be the most interesting one, particularly for the critical separation of heroin, papaverine, and acetylcodeine. Thus, this material was selected to reach the optimal separation.

In conclusion, the best separation of the 12 analytes of interest in UHPLC was achieved with an Acquity BEH Shield RP18 50×2.1 mm, 1.7-μm column operating at 30 °C, with a flow rate of 500 μL/min. The optimal gradient found with HPLC optimization software varies from 27% to 69.5% MeOH in 3.48 min, followed by a step at 69.5% for 0.52 min and was carried out at pH 9. For MS detection, dwell times for each SRM transition were adjusted in agreement with the required sensitivity. As the response was sufficient for most of the compounds, the lowest possible dwell time, 5 ms, was selected for most of the analytes. On the other hand, dwell times of 40 and 50 ms were selected for meconine and acethylthebaol, respectively, because these two compounds provided the lowest signal-to-noise ratio. Using this strategy, it was possible to maintain an acquisition rate of 5 Hz and thus obtain reliable data even with thin UHPLC peaks (at least 15–20 data points across the peaks).

Compared with the original GC-MS separation carried out in around 23 min (Fig. 1), the optimized UHPLC–MS/MS strategy allows a reduction of analysis time down to only 4 min (Fig. 5), while the minimal resolution remains at least equal to 1.5 in UHPLC–MS/MS (lower resolution observed for the critical pair, 6-MAM/procaine). Two

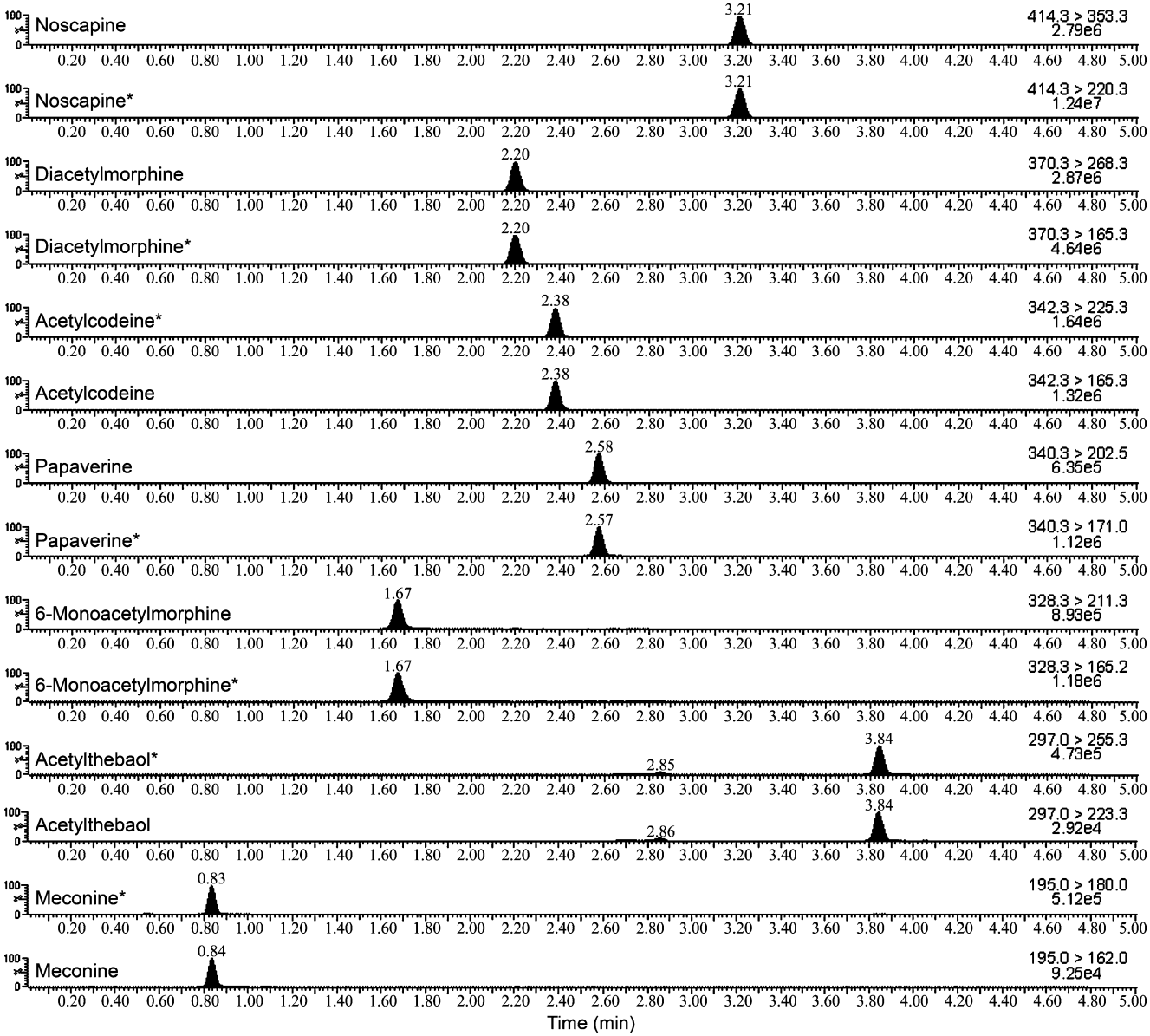

Fig. 6 Extracted ion chromatogram of the UHPLC–MS/MS separation of an unknown heroin. Two transitions of each of the seven major alkaloids of heroin were followed by MS/MS using the conditions described in Table 3. *Star superscript* indicates the transisitions used for peak integration. Analytical conditions were similar to Fig. 5

transitions by compounds were recorded to increase selectivity (Fig. 6), and the most intense transition was selected to integrate the peaks and conduct the transfer. In addition, the absence of sample derivatization prior to the analysis leads to conclude that UHPLC–MS/MS represents the fastest analytical approach for heroin profiling.

Transfer results

Table 4 summarizes the results obtained for the transfer of UHPLC–MS/MS data to the GC-MS database. For a given tested model, adjusted R^2 and Q^2 of the seven MLRs (one MLR by compound) were averaged to estimate the overall predictive abilities of each model. In fact, the frequency of successful transfer of the validation set is much more representative of the predictive abilities of each model than for the calibration set which is used for the establishment of the transfer models (B.5.). Most models offered relatively low predictive abilities and some of them were overfitted. Indeed, the observed differences between the frequency of successful transfer for the calibration (e.g., 86.1% for the quadratic model) and validation sets (e.g., 68.2% for the quadratic model) clearly illustrated this problem. A correlation can be observed between the mean adjusted R^2 and the calibration set frequency of successful transfer. This means that if the model was well adjusted on the calibration

Table 4 Model mean adjusted coefficients of determination (adjusted R^2), mean coefficient of prediction (Q^2), and frequency of succesful transfer obtained for the calibration set and the validation set

Model name	Calibration set (36 samples)			Validation set (22 samples)
	Mean adjusted R^2	Mean Q^2	Frequency of succesfull transfer (%)	Frequency of succesfull transfer (%)
Linear	0.695	0.627	80.6	81.8
Quadratic	0.732	0.671	86.1	68.2
Cubic	0.728	0.646	86.1	68.2
Poly-linear	0.802	0.750	94.4	81.8
Poly-quadratic	0.871	0.842	97.2	68.2
Complete	0.971	0.902	100.0	63.6
Simplified	0.803	0.725	97.2	95.5

points (relatively high adjusted R^2), it can be assumed that the "GC-like" samples coming from the calibration points will be successfully transferred. However, no correlation was observed between mean Q^2 and the validation set frequency of successful transfer. Linear and poly-linear models gave reasonably good frequencies of successful transfer and respectively offered better results than quadratic and poly-quadratic models. In this case, the addition of quadratic (or cubic) terms in transfer models did not increase the frequency of successful transfer. The frequency of successful transfer directly depends on the height at which dendrograms were cut (see "Significant clustering height calculation" and "Evaluation of transfer model predictive abilities"). As the frequency quickly increased when the cutting height was slightly augmented, high frequencies of successful transfer were obtained with relatively low cutting height.

However, it is important to keep in mind that the ionization sources of the two analytical techniques were very different: ESI operating at atmospheric pressure (i.e., soft ionization mode) and EI in vacuum conditions (i.e., hard ionization mode) for UHPLC–MS/MS and GC-MS, respectively. Despite this dissimilarity, the simplified model, which includes the mathematical terms allowing the maximization of the adjusted R^2 (mainly linear terms), was able to accurately predict the samples included in the validation set (see Table 4).

Furthermore, the current size of the used sets already allowed a very good preliminary estimation of model predictive ability even if increasing the size of calibration set and especially validation set could also allow a better evaluation of their abilities. In addition, simplified models were uncomplicated to manage as only three terms were generally selected during the Pareto analysis. These seven simplified models (one for each compound) offered a probability of 95.5% of successful transfer for future heroin samples.

Conclusion

The transfer of results from an analytical technique to another is a delicate task that requires paying special attention to the physical and chemical phenomena that regulate analytical signals obtained and to the statistical techniques employed. Indeed, the transfer complexity depends on the similarity of data nature and, particularly in the case of MS detection, on the ionization sources. In this context, simplified models with an average of three mathematical terms have achieved very promising results (i.e., almost totally successful transfer), paving the way for the methods to be implemented for this type of transfer.

GC-MS currently remains the technique of choice to achieve heroin profiling, mainly because this technique is generally present in most analytical chemistry, toxicology, and forensic science laboratories. However, UHPLC–MS/MS offers enormous benefit to avoid the derivatization step, to move forward to the trend of "fast forensic," and easily achieve quantification of investigated alkaloids. In light of the results, the transfer of heroin profiling to UHPLC–MS/MS could thus be considered in the near future.

In conclusion, a growing interest could be paid to explore other pretreatments and mathematical models while evaluating the additional profit they could bring in the transfer of analytical techniques. Until this potential improvement, the sequential use of PCA and HCA has already been proven to be a simple and powerful unsupervised statistical tool that achieves meaningful classifications (data-driven clustering) with respect to the data set studied in this work. The present transfer strategy can therefore be regarded as a very promising starting point for future statistical methodologies in the framework of the transfer of other data sets (e.g., cocaine and other illicit drugs) and/or new analytical techniques.

Acknowledgments The author would like to thank the Fonds de la Recherche Scientifique (F.R.S.-FNRS) for the research grant and the Walloon Region of Belgium for the FIRST-DEI convention funds no. 516130. The authors wish also to acknowledge Waters for the loan of ACQUITY TQD instrument as well as Dr. A. Tei from Waters for the implementation of the APPI source.

Appendix

The fully described equations of the seven simplified models are given here.

$$MEC_{GC} = -0.151 + 0.990\,MEC_{UHPLC} + 0.056\,AC_{UHPLC}$$
$$AC_{GC} = 0.217 + 1.087\,AC_{UHPLC} - 0.332\,PAP_{UHPLC}$$
$$AcTB_{GC} = 0.264 + 1.129\,AcTB_{UHPLC} + 0.176\,PAP_{UHPLC} - 0.069\,AcTB_{UHPLC} \cdot MAM_{UHPLC} - 0.074\,AcTB_{UHPLC}^{2} \cdot PAP_{UHPLC}$$
$$MAM_{GC} = 0.136 + 0.882\,MAM_{UHPLC} - 0.234\,AcTB_{UHPLC}$$
$$DAM_{GC} = 0.246 + 0.546\,DAM_{UHPLC} + 0.442\,AC_{UHPLC} - 0.261\,MAM_{UHPLC}$$
$$PAP_{GC} = -0.607 + 0.207\,PAP_{UHPLC} + 0.289\,NOS_{UHPLC} - 0.034\,MEC_{UHPLC}^{2} + 0.225\,PAP_{UHPLC} \cdot MAM_{UHPLC}$$
$$NOS_{GC} = -0.359 + 0.729\,NOS_{UHPLC} - 0.220\,AC_{UHPLC} + 0.315\,NOS_{UHPLC} \cdot PAP_{UHPLC}$$

Mathematical terms were labeled according to the compound, with subscripts corresponding to the analytical technique.

References

1. Guéniat O, Esseiva P (2005) Heroin and cocaine profiling, 1st edn. Presses polytechniques et universitaires romandes, Lausanne
2. Esseiva P, Dujourdy L, Anglada F, Taroni F, Margot P (2003) Forensic Sci Int 132:139–152
3. Dujourdy L, Barbati G, Taroni F, Guéniat O, Esseiva P, Anglada F, Margot P (2003) Forensic Sci Int 131:171–183
4. Hibbert DB, Blackmore D, Li J, Ebrahimi D, Collins M, Vujic S, Gavoyannis P (2010) Anal Bioanal Chem 396:765–773
5. Lurie IS, Hays PA, Garcia AE, Panicker S (2004) J Chromatogr A 1034:227–235
6. Morello DR, Cooper SD, Panicker S, Casale JF (2010) J Forensic Sci 55:42–49
7. Morello DR, Meyers RP (1995) J Forensic Sci 40:957–963
8. Perillo BA, Klein RFX, Franzosa ES (1994) Forensic Sci Int 69:1–6
9. Esseiva P, Anglada F, Dujourdy L, Taroni F, Margot P, Du Pasquier E, Dawson M, Roux C, Doble P (2005) Talanta 67:360–367
10. MacNair JE, Lewis KC, Jorgenson JW (1997) Anal Chem 69:983–989
11. Swartz ME (2004) LabPlus Int 18:6–9
12. Kyranos JN, Cai H, Wei D, Goetzinger WK (2001) Curr Opin Biotechnol 12:105–111
13. Joseph M (2005) LC GC N Am 23:3–34
14. Lurie IS, Toske SG (2008) J Chromatogr A 1188:322–326
15. Frank IE, Todeschini R (1994) The data analysis handbook. Elsevier, Amsterdam
16. Stella C, Rudaz S, Gauvrit J-Y, Lantéri P, Huteau A, Tchapla A, Veuthey J-L (2007) J Pharm Biomed Anal 43:89–98
17. Boccard J, Grata E, Thiocone A, Gauvrit J-Y, Lantéri P, Carrupt P-A, Wolfender J-L, Rudaz S (2007) Chemom Intell Lab Syst 86:189–197
18. Ward JH (1963) J Am Stat Assoc 58:236–244
19. Suzuki R, Shimodaira H (2006) Bioinformatics 22:1540–1542
20. Guillarme D, Ruta J, Rudaz S, Veuthey J-L (2010) Anal Bioanal Chem 397:1069–1082
21. Guillarme D, Schappler J, Rudaz S, Veuthey J-L (2010) TrAC-Trend Anal Chem 29:15–27
22. Marchi I, Rudaz S, Veuthey J-L (2009) Talanta 78:1–18
23. Cai SS, Syage JA, Hanold KA, Balogh MP (2009) Anal Chem 81:2123–2128
24. Schappler J, Guillarme D, Prat J, Veuthey J-L, Rudaz S (2007) Electrophoresis 28:3078–3087
25. Bones J, Macka M, Paull B (2007) Analyst 132:208–217
26. Rook EJ, Hillebrand MJX, Rosing H, van Ree JM, Beijnen JH (2005) J Chromatogr B 824:213–221
27. Berg T, Lundanes E, Christophersen AS, Strand DH (2009) J Chromatogr B 877:421–432
28. Schappler J, Nicoli R, Nguyen D, Rudaz S, Veuthey J-L, Guillarme D (2009) Talanta 78:377–387
29. Nguyen D, Guillarme D, Heinisch S, Barrioulet MP, Rocca J-L, Rudaz S, Veuthey J-L (2007) J Chromatogr A 1167:76–84
30. Majors RE (2009) LC GC N Am 27:458–471